
Chapter 3: Guide to the Data

This section provides some practical information about how NLS variables are collected, created, and arranged on the data set along with accompanying hard copy and electronic documentation. NLS variables are derived, for the most part, directly from survey instruments; arranged both numerically and topically within the NLS documentation system; and presented, within a codebook, along with full and complete information on each variable.

The first section of this chapter describes the different survey instruments used to collect the raw NLSY79 data during the field period. This section also explains how question numbers have been assigned during the various survey years. Next, the guide to the data discusses the primary types of NLS variables and the process by which each is assigned a reference number and title that serve to identify it throughout the NLS documentation system. The third section reviews the codebook, the information about each variable contained on the data set, and the accompanying paper documentation. This discussion will help users understand how to interpret the various pieces of data presented in the NLS documentation system. Finally, this chapter gives researchers some basic instruction on using the search functions to find variables relating to the area of interest.

User Notes: The focus of these sections will be on accessing variables found on what are called the main (or geocode) data files for the NLSY79 and NLSY79 Child data files.

3.1 Survey Instruments

The primary variables found within the main data set are derived directly from one or more survey instruments, e.g., questionnaires, household interview forms, etc. This section describes each of the NLSY79 survey instruments in the order that they appear within Table 3.1.1. It also explains the conventions used in the NLSY79 documentation system to identify questionnaire items from some of the primary survey instruments. An additional document, the interviewer reference manual, provides background information on specific survey instruments. While not actually a survey instrument, this document is also described within this section.

Table 3.1.1 Types of NLSY79 Survey Instruments & User Aids

1978 Household Screener, Household Interview Forms	Children of the NLSY79 Supplements
Interviewing Aids	Interviewing Aid
<i>Face Sheet</i>	<i>Child Face Sheet</i>
<i>Information Sheet</i>	Questionnaire Supplements
<i>Children's Record Forms (CRF)</i>	<i>Child Supplement (CS)</i>
Questionnaires	<i>Mother Supplement (MS)</i>
Questionnaire Supplements	<i>Child Self-Administered Supplement (CSAS)</i>
1979 High School Survey	<i>Child Self-Administered Supplement</i>
1980–83 Transcript Surveys	<i>Confidential Card</i>
1980 <i>Illegal Activities Form J</i>	<i>Young Adult Survey</i>
<i>Employer Supplements (ES)</i>	<i>Young Adult Self-Report Booklet (YASRB)</i> ¹
1983 <i>Fertility Supplement</i>	1995 School Survey
<i>Confidential Abortion Forms</i>	Interviewer Reference Manuals (Q by Qs) and CAPI Help
1988, 1992, 1994 & 1998 <i>Drug Use Supplement</i>	Information
1988 <i>Childhood Residence Calendar</i>	

¹ Used through 1998, in 2000 fielding, retained questions were incorporated into the CAPI instrument.

Questionnaire Item or Question Number: This generic term refers the user to the printed source of data for a given variable. A questionnaire item may be a question, a check item, or an interviewer's reference item that appears within one of the survey instruments. Each questionnaire item has been assigned a number or a combination of numbers and letters within the NLSY79 documentation system to assist the user in linking each variable to its location in a survey instrument.

NLSY79 questionnaire item assignment is complex and varies across survey years and instruments. For some years, NLSY79 questionnaire item identification is dependent upon various combinations of the deck and column numbers used in data entry that are printed to the right of the answer categories on the survey instrument. In other years, designation is made by section and question numbers. Specific information on the conventions used appears below, after each relevant instrument, under the subheadings "Question Numbering."

NLSY79 Survey Instruments

A unique set of survey instruments has been used during each survey year to collect information from respondents. The term "survey instrument" is used to refer to (1) the questionnaires that serve as the primary source of information on a given respondent; (2) questionnaire supplements fielded during select survey years that contain additional sets of questions; and (3) documents such as the household interview forms or household record cards that collect information on members of each respondent's household.

Users should be aware that, while the source of the majority of variables in the main NLSY79 data files is the questionnaire or one of the other survey instruments, certain NLSY79 variables are created either from other NLSY79 variables or from information found in an external data source (see “Types of NLSY79 Variables” later in this chapter).

Household Interview Forms

Each NLSY79 interview includes the collection of information on the members of each respondent’s household. For NLSY79 respondents, such household data are collected prior to the administration of the main questionnaire and for many years used separate survey instruments called the *Household Interview Forms*. Both the instruments used for the yearly household data collection and the household screening instruments that were used to draw the samples of respondents are described below.

NLSY79 1978 Household Screener and Interviewer’s Reference Manual: This document (fully titled *NLSY-National Longitudinal Survey of Labor Force Behavior Interviewer’s Manual-Household Screening, NORC 1978*) contains detailed information on the 1978 screening of households conducted by NORC from which the civilian youth samples (the cross-sectional and supplemental samples) were drawn. It provides a copy of the short 25-question screener, question-by-question specifications for administering the form, and a sample completed screener. Most of the information collected on each respondent during the screening is presented within the data set. The screener is the source for important data such as the sex and race/ethnicity variables that were used to assign each respondent to a specific NLSY79 subsample, as well as the relationship codes (e.g., brother, sister, husband, wife) that allow researchers to identify related NLSY79 respondents who shared a household at the time of the screening.

Question Numbering: Question numbers for the 1978 screener were arbitrarily assigned by NORC using an artificial questionnaire section number that followed the last section of the 1979 questionnaire (“Section 25” for all screener variables) even though the actual administration of the screener preceded that of the 1979 questionnaire.

Users should note that screener questions are identified within the documentation as 1979 variables even though these data were collected during 1978. Most variables from the screener use the phrase HOUSEHOLD SCREENER at the beginning of the variable title, appear physically within the codebook after the 1979 household record series, and have been placed within the “Misc. 1979” area of interest.

Household Interview Forms: Yearly household information for the NLSY79 is collected from either the respondent or the head of household prior to the administration of the main questionnaire. NLSY79 *Household Interview Forms* are used to (1) enumerate all persons currently living in the respondent’s household; (2) record information about each person’s age, highest grade completed, work experience in

the past year, and relationship to the respondent; and (3) collect, during the 1979–86 surveys, certain family income information. Information on household members is collected using the questions on the *Household Interview Forms*; however, much of the information is actually recorded on the “Household Enumeration” section of the *Face Sheet* discussed below.

During the 1979–86 interviews, different versions of the *Household Interview Forms* were administered depending upon the type of residence of the respondent. Version A was used if the respondent was living with his/her parents (or in-laws), in which case the interview was conducted with the respondent’s parents (or in-laws) in order to gather information on household income sources. Version B was used if the respondent was living in group quarters, such as a dormitory or the military, or in temporary facilities, such as a hospital or prison, and was administered to the respondent. If the respondent had a permanent residence elsewhere, the household interview gathered information about that household. Version C was administered to the respondent if s/he was living in his/her own dwelling unit, military family housing, an orphanage, a religious institution, or other individual quarters or was the head of a family unit. Table 4.20.1 in the “Household Composition” section of this guide depicts, by survey year, the universe and residential unit(s) specific to each form.

During the first eight survey rounds, many respondents were younger than 18 and living with their parents; thus, Version A was frequently used. Beginning with the 1987 survey, all respondents were 21 or older and living predominantly on their own; consequently, the household interview forms were consolidated into a single version. For 1979–86, these forms appear as separate documents. Beginning with the 1987 interview, household interview questions were incorporated within each year’s questionnaire. Some variation in administration of these forms has occurred over survey years. Users should refer to each survey year’s *Interviewer’s Reference Manual* for more information.

Interviewing Aids

Certain instruments used during fielding of the NLSY79 and NLSY97 child surveys provide researchers with interview-, respondent-, and child-specific information that appears as variables within the NLSY79 data files.

Face Sheet: Immediately prior to fielding, a *Face Sheet* is computer-generated for each respondent and forwarded to the interviewer assigned to that case. The *Face Sheet* contains (1) various items of respondent-specific information (name, address, phone number); (2) information about each member of the household or family unit as of the last interview (full name, sex, relationship to youth, education, and whether the household member worked during the year), generated from the most recent administration of the *Household Interview Forms*; (3) a historical overview of previous interview rounds (whether the respondent refused to be interviewed, the case was converted [i.e., the respondent was interviewed after

initially refusing], the interview was complete or incomplete, etc.); and (4) for the 1980–86 survey years, information on the version of the *Household Interview Form* that was used in the previous interview. This information is used to alert the interviewer and field manager to potential problems, assist them in preparing a successful location and fielding strategy, and provide details necessary to conduct an efficient interview, e.g., a listing of previous employers. Information about the respondent’s household and family unit from each survey year’s *Face Sheet* appears as a set of variables in the “Household Record” area of interest on the NLSY79 main data set. Sample *Face Sheets* for most survey years can be found in the various *Interviewer Reference Manuals*.

Information Sheet: This document contains data on the respondent from the previous interview that will be referred to and used to update information during the interviewing process. Items found on this document include marital status, high school completion status, university last attended, names of previous employers, training program enrollment, and pregnancy status. This information enables the interviewer to accurately route the respondent through the relevant sections of the questionnaire and provides on-the-spot reconciliation of earlier errors. *Information Sheet* items appear within the NLSY79 data set (“Last Interview Information” area of interest). Beginning with the 1993 interviews, the information sheet is incorporated into the CAPI instrument. Sample *Information Sheets* can be found in the *Interviewer Reference Manuals*. In CAPI surveys, information sheet data are stored electronically on the interviewer’s laptop and accessed by the survey program during the interview; no paper information sheet is used.

Children’s Record Forms (CRF) (1985–92): This interviewing aid containing information on biological (collected each survey) and nonbiological (i.e., adopted or step-; collected biennially) children was used in the 1985-92 survey to (1) provide identification numbers, names, dates of birth, sex, and deceased/adopted status for each child and (2) identify special sections of the main questionnaire (e.g., immunization, feeding, etc.) that needed to be administered for particular children. Sample *Children’s Record Forms* can be found in the *Interviewer’s Reference Manuals*. Beginning with the 1993 interviews, this form is incorporated into the CAPI instrument. As with information sheets, these data are automatically accessed by the survey program during CAPI interviews, so the hard copy CRF is no longer needed.

Questionnaires

There are separate and distinctly different questionnaires for each survey year of the NLSY79 and the NLSY79 children. Each questionnaire is organized around a set of topical subjects, the titles of which usually appear on either the first page of each section of the questionnaire or as a header.

The questionnaires are critical elements of the NLSY79 documentation system and should be used by each researcher to ascertain the wording of questions, coding categories, and the universe of respondents asked to respond to a given question.

NLSY79 questionnaires record (1) interview dates; (2) responses to the topical survey questions (see discussion below); (3) locating information which will assist NORC in finding the respondent for the next interview; and (4) interviewer remarks on such topics as the race and sex of respondent, language in which the interview was conducted, interviewer's impressions, etc. *Show Cards*, interviewing aids used in conjunction with the questionnaire, list the various possible response categories for selected questions and help the respondent keep the more complicated response categories in mind.

NLSY79 questionnaires explore the following core topics: current labor force status, jobs and employers, work experience and attitudes, training, assets and income, family background, marital history, fertility, regular schooling, military service, and health. Additional sets of questions on such topics as childcare, alcohol use, drug use, job search methods, educational/occupational aspirations, school discipline, pre- and post-natal health behaviors, delinquency, childhood residences, and so forth have been fielded during select survey years.

During the 1979–92 paper-and-pencil (PAPI) interviews, questionnaires and other survey instruments were preprinted paper products used during fielding. With the advent of computer-assisted interviewing (CAPI) in 1993, the “questionnaire” became a series of visual screens that not only told the interviewers what questions to ask but provided helpful instructions on how to administer the interview. Separate supplemental documents such as the job-specific *Employer Supplements* were integrated into the electronic main questionnaire. Generation of a hard copy document became a post-survey process. NLSY79 CAPI questionnaires incorporate some helpful elements of the traditional codebook, with reference numbers assigned to variables and greater specificity on coding and universes provided within each codeblock.

Question Numbering: The conventions used to assign question numbers within the NLSY79 documentation system vary by survey year and are based on various combinations of the questionnaire section number, the question number, and/or the deck and column numbers (Table 3.1.2). Users can locate a variable within the hard copy codebook—which represents each question fielded in the same order as it appears within the questionnaire—by finding the question number which appears (in parentheses) to the right of each reference number.

Table 3.1.2 NLSY79 Question Numbering Conventions

Survey Year	Designated By	Example
1979	Section # (S) & Question # (Q)	S02Q01: Question 1 in Section 2
1980–82	Section # (S), Deck # (D) & Column #	S06D1314: Question appearing in Section 6, deck 13, column 14
1983–87, 1989–92	Deck # & Column #	Q0413: Question appearing in deck 4, column 13
1988	Section # & Question # (Q)	Q5.3: Question 3 in Section 5
1993– 2002	Section #, Question # (Q) & Loop # as applicable	Q5-26.3: Question 26 in Section 5, with the appended .03 representing the third loop

Deck and column numbers are vestigial items that were used to locate the data when it was input on punch cards. The deck numbers are printed at the upper right hand corner of each page in the survey instruments and at the beginning point for each new deck for the 1980 through 1992 instruments. The column numbers are printed to the left of the response categories. If the variable contains more than one digit, the column reference is to the starting column for that variable.

User Notes: Although NLSY79 questionnaires are to some extent topically arranged, the user should be aware that the absence of a section title on a given subject does not mean that no questions on that topic were fielded during that survey year. For example, the 1987 and 1989 NLSY79 questionnaires contain no section entitled “Childcare”; however, a small number of childcare questions were asked in those years and appear within the “Fertility” section of the questionnaires.

Questionnaire Supplements

Separate instruments called “supplements” have been used since the onset of the NLSY79 to administer distinct sets of questions. The NLSY79 has made extensive use of supplements for collecting information from separate universes such as schools or children or for administering confidential sets of questions on illegal activities or abortion. The following section describes each supplemental instrument used for the NLSY79. The use of such separate supplements has diminished with CAPI-administered interviews. In the main youth and young adult instruments, all supplements are now incorporated as electronic modules in a questionnaire. Children still use multiple supplements, one self-report, one interviewer-administered, and one completed by the mother.

Illegal Activities Form J (1980): This confidential questionnaire supplement, administered during the 1980 survey, contains a series of questions designed to collect information on the extent of respondents’ participation in various delinquent and criminal activities such as skipping school, alcohol/marijuana use,

vandalism, shoplifting, drug dealing, and robbery. This series supplements those on reported contacts with the criminal justice system collected within the main questionnaire.

Employer Supplement: Information about each employer for whom a NLSY79 respondent has worked since the last interview has been collected since 1980. One *Employer Supplement* is administered for each employer and contains questions about gaps when the respondent was not working, the number of hours worked, the type of work done, and the wages earned at that job. Note: Comparable information for the 1979 survey can be found in the “On Jobs” section of the main questionnaire and within the separate single sheet 1979 *Employer Flap*. Beginning with the 1993 CAPI interviews, all employer supplement questions appear within the body of the main questionnaire.

Question Numbering: Five numbering systems have been used to identify questionnaire items within the *Employer Supplement* (Table 3.1.3). Although data from up to 10 jobs are collected, the main data set includes information on only the first five jobs since few individuals work at more than five jobs between interviews. Data on all ten jobs are used to construct a series of summary variables for hours and weeks worked; see the “Labor Force Status,” “Time & Tenure,” and “Work Experience” sections for more information.

Table 3.1.3 *Employer Supplement* Question Numbering Conventions: 1980–2002

1980–87 1989–91	A supplement identifier, i.e., the letter B, representing the first supplement, through F, the fifth supplement, is combined with the deck and column numbers preprinted in the instrument. The deck numbers for the first <i>Employer Supplement</i> would be B1, B2, B3, and B4 while the second supplement would use C with each deck and column number. The question number QB140 thus refers to B (the first supplement), 1 (deck 1), 40 (column 40), while QC166 refers to <i>Employer Supplement C</i> , deck 1, column 66.
1988	Letter designations, i.e., ESB, ESC, ESD, ESE, ESF, continue to identify the specific supplement in use; however, deck and column numbers are not used. Appended to the supplement identifier is the actual question number as printed in the supplement. For example, ESB.1 refers to the first supplement, question 1.
1992	A series of supplemental deck numbers are attached to the column numbers preprinted in the supplement. Question numbers 7439–7831 refer to information collected in the first supplement, 7939–8331 to the second supplement, 8439–8831 to the third supplement, 8939–9331 to the fourth supplement, and 9439–9831 to the fifth supplement.
1993–1996	The designation QES and a number, e.g., QES5, indicates that this series of questions collected information about the fifth employer. Hyphenated numbers attached to the QES5, e.g., QES5-26, QES5-27, etc. indicate the specific question number within the series, while a decimal number following a question number, QES5-26.3, reflects the third repetition of that question for that employer.
1998–2002	Beginning in 1998, the number identifying the employer was moved to a decimal after the question number. The question previously labeled QES5-26.3, for example, was now designated as QES-26.05.03. The decimal number “.05” indicates this information was collected about the fifth employer. Again, “.03” represents the third repetition of question 26.

Fertility Supplement (1983): Respondents (both male and female) who were not interviewed during 1982 were administered a special set of supplementary fertility questions during the 1983 survey. The *Fertility Supplement* was designed to collect complete fertility data, including all live births for males and females, and all pregnancy losses and contraception between pregnancies for females. For those not interviewed in 1982, these questions replaced the fertility questions found in Section 10 of the 1983 questionnaire.

Confidential Abortion Forms: Biennially beginning in 1984, female NLSY79 respondents have completed a short confidential abortion form which elicited information on the number and dates of each abortion. Copies of these supplementary questions are provided within the survey instrument sets. The 1984 form also collected information on the dates that respondents left school prior to 1979 if leaving school was associated with early childbearing.

Drug Use Supplement (1988, 1992, 1994 & 1998): The 1988 supplement contains the confidential set of drug use questions which were, through a random assignment process, self-administered by the respondent in half of the cases and administered by the interviewer in the other half. Questions were asked on age at first use of marijuana and cocaine, extent of lifetime and most recent use, and method(s) practiced in using cocaine. The 1992 and 1994 supplements contain the confidential set of questions on respondents' use of cigarettes, alcohol, marijuana, cocaine, or other drugs. Users should note that while the 1988 and 1992 supplements are bound as separate booklets, the 1994 and 1998 supplements are bound with the main questionnaire.

Childhood Residence Calendar (1988): The 1988 questionnaire contained a special section detailing the living arrangements of respondents from birth through age 18. The *Childhood Residence Calendar*, the interviewing aid used to collect these data, depicts for each year of life the type of parent (biological-, adoptive-, or step-) with whom each respondent lived for at least four months and, for those ages when s/he was not living with a parent, in what other arrangements the respondent resided, such as, with grandparents, foster parents, friends, or in a children's home, detention center, or other institution.

Supplemental Data Collections

High School Survey (1980): A supplemental survey of the last secondary school attended by civilian NLSY79 respondents was conducted in 1980. This survey gathered information on each school's grading system, course offerings, dropout rate, student body composition, and faculty characteristics, as well as respondent scores from a variety of intelligence and aptitude tests. Copies of the high school survey instruments, the "School Questionnaire" and the "Student's School Record Information" form, are included within the documentation item called the *NLSY High School Transcript Survey: Overview and Documentation*.

Transcript Surveys (1980–83): Transcript information on up to 64 courses was collected from high school records for civilian NLSY79 respondents who were expected to complete high school within the United States. A copy of the instrument used to collect transcript information, called the “Transcript Coding Sheet,” is included within the *NLSY High School Transcript Survey: Overview and Documentation*.

ASVAB: The *Armed Services Vocational Aptitude Battery (ASVAB)* was administered to most NLSY79 respondents in 1980 as part of a Department of Defense effort to renorm this military enlistment test. The scores from this supplemental data collection are included in the NLSY79 data file. For details, see the “Aptitude, Achievement & Intelligence Scores” section of this guide.

Children of the NLSY79 Interviewing Aids

Child Face Sheet: This interviewer information sheet, used during the 1988–98 fieldings, contained information on the child’s ID, name, mother’s sample type (1990 only), *Child Supplement* interview date, child’s date of birth, child’s age at date of child supplement, PPVT age, school grade, whether child has had menses, interviewer ID, and a grid indicating which assessments should be administered.

Young Adult Face Sheet: This interviewer information sheet used during the 1994–2002 fieldings, contained ID, name, all known locating data, mother’s address, and last interview date for the young adult and mother.

Table 3.1.4 Contents of the Separate Child Instruments

<i>Child Supplement (CS)</i>	<i>Mother Supplement (MS)</i>	<i>Child Self-Administered Supplement (CSAS)</i>	<i>Young Adult Survey and Young Adult Self-Report Booklet (YASRB)¹</i>
Background Child Health Body Parts Memory for Location Verbal Memory Self-Perception Profile for Children (SPPC)/ What I Am Like Digit Span PIAT Math & Reading PPVT Interviewer evaluation of testing conditions Interviewer observations of home environment	HOME – Short Form How My Child Usually Acts Motor & Social Development Behavior Problems Index School & Family Background Interviewer Remarks	Interactions with Parents TV Viewing Computer Use & Attitudes Responsibilities in the Home Time Use after School School Satisfaction Religion Peer Relationships Dating; Marriage & Childbearing Expectations Sexual Activity & Sex Education Substance Use Work for Pay Interviewer Remarks	Household Record Family Background Dating/Marital History Schooling & Military Jobs & Training Fertility Health Childcare Income & Assets Attitudes Drug & Alcohol Usage Sexual Activity & Abortion Delinquent & Criminal Activity

¹As of 2000, the YASRB was integrated into the CAPI survey.

Children of the NLSY79 Questionnaire Supplements

Variables in the NLSY79 child data set are derived from three sources; questionnaire supplements administered in even numbered years since 1986, the child cognitive and socio-emotional assessments administered during the same surveys, and relevant mother and child info taken from the main NLSY79 survey. The special child supplements and the main topics covered in each are listed in Table 3.1.4.

Child, Mother & Child Self-Administered Supplements: Survey schedules called the *Mother Supplements (MS)*, the *Child Supplements (CS)*, and the *Child Self-Administered Supplements (CSAS)* are used to (1) administer a battery of cognitive and socio-emotional assessments to children born to female NLSY79 respondents; (2) gather information from each mother on her children's home environment, behavior, health, and motor-social-cognitive development; (3) collect information from children ages 10 years and over on a wide range of topics; and (4) collect from interviewers their evaluation of the testing conditions and observations of the child's home environment. Abbreviated versions of the mother and child supplements called the *Infant Supplement* and the *Mother of Infant Supplement* were used in the 1986 survey for households in which only children under eight months of age resided.

Contents of the *Child Supplement*, *Mother Supplement*, and *Child Self-Administered Supplement* differ somewhat across survey years but typically consist of the sections depicted in Table 3.1.4. Specific content listings for the *Mother Supplement* can be found either on the inside front or back cover of the instrument; accompanying charts indicate the specific assessments administered for children of varying ages.

Child Self-Administered Supplement Confidential Card: Starting with the 1988 survey, a separate confidential form, administered to children age 13 and over, collected information on whether the child had ever engaged in sexual intercourse and, if so, age and date of first intercourse. The 1992, 1994, 1996, and 1998 versions include the dates of any live births.

Young Adult Survey and Young Adult Self-Report Booklet: Beginning in 1994, a new survey entitled the *NLSY79 Young Adult Survey* was introduced. This survey was designed to ask NLSY79 style questions to the oldest children of NLSY79 mothers. Children who will be age 15 or older by the end of a given survey year complete the young adult instruments rather than the child instruments described above. The CAPI questionnaire gathers information on each young adult's experiences. A separate *Young Adult Self-Report Booklet* asked respondents sensitive questions about drug and alcohol usage as well as criminal and sexual activity. In 1998, the Young Adult group was capped at age 20. In 2000, the primary mode of administration for the Young Adults was telephone, so the self-report booklet was incorporated into the main survey instrument. Young Adults interviewed in person were given the option of self-administration using the interviewer's laptop.

Question Numbering: Variables on the NLSY79 child data set are derived from three sources: the questionnaire supplements described above, the child cognitive and socio-emotional assessments administered at the time of the survey, and relevant mother and child information taken from the main NLSY79 interview.

Questions from the specific mother and child assessment instruments are identified by “MS” for *Mother Supplement* or “CS” for *Child Supplement*. The two digits that follow this prefix refer to the year of the survey. Through 1992, the next two numbers refer to the deck number printed at the top of each page in the survey instrument, and finally to the column number printed alongside each question in the survey instrument. For instance, question CS861217 refers to deck 12, column 17 of the 1986 *Child Supplement*. Deck numbers for the *Child Self-Administered Supplement* follow consecutively those in the regular *Child Supplement*; questions are numbered in the same manner as the regular *Child Supplement*. Starting in 1994, question numbers more closely parallel those of the main youth CAPI instrument. However, the assessment items from the *Child Supplement* and constructed variables in the file have been assigned mnemonic variable names.

The young adult question numbering conventions parallel those of the main youth CAPI questionnaires (1993, 1994, 1996, 1998, 2000, and 2002). Each question number indicates the section, question within that section, and loop number if applicable. As of 1998, loop number is no longer part of the question name because loops are dynamic. For more information, refer to the discussion of question numbering for the main youth earlier in this section.

1995 School Survey: In 1995 a separate child school survey was conducted for children born to female respondents of the NLSY79. The first part of this data collection was a questionnaire addressed to the school principal that solicited a variety of information about characteristics of the school and school policies. Second, the school central staff was asked to complete a questionnaire that focused on the individual youth’s academic success, social adjustment, and involvement in a variety of school activities. This questionnaire also gathered selected characteristics and policies relevant to the grade level in which the youth was enrolled. In addition, the youth’s transcript was obtained whenever possible.

Interviewer’s Reference Manual (Question-by-Question [Q by Q] Specifications)

Each questionnaire or set of survey instruments is accompanied by an *Interviewer’s Reference Manual*. This document provides NORC interviewers with background information on the NLSY79 and detailed question-by-question instructions for administering and coding the questionnaire, *Employer Supplement*, *Household Interview Forms*, and other survey supplements. Separate *Q by Q*’s exist for each NLSY79 survey year as well as for the NLSY79 children’s *Child Supplement* and *Mother Supplement*. Printed copies

of the CAPI help screen information, which each interviewer could access during the course of the interview, replace the traditional interviewer's manual instrument beginning with the 1993 release.

3.2 Types of Variables

There are six types of variables present in the NLSY79 data. Some are the raw answers provided by the respondent, while others are constructed. The type of variable impacts: (1) the title or variable description naming each variable; (2) the physical placement of each variable within the codebook; and (3), the location of a variable within a given area of interest. Types of variables include:

- (1) Direct (or raw) responses from a questionnaire or other survey instrument.
- (2) Edited variables constructed from raw data according to consistent and detailed sets of procedures, e.g., occupational codings, *KEY* variables, etc.
- (3) Constructed variables based on responses to more than one data item, either cross-sectionally or longitudinally, and edited for consistency where necessary, e.g., variables on the NLSY79 Supplemental Fertility File ("Fertility and Relationship History/Created" area of interest).
- (4) Constructed variables from data provided on a non-NLS data set, e.g., the *County & City Data Book* information present on the NLSY79 geocode data files.
- (5) Variables provided by NORC or another outside organization based on sources not directly available to the user, e.g., the high school survey and transcript data, scores from the *Armed Services Vocational Aptitude Battery*, etc.
- (6) Data collected from or about one universe of respondents reconstructed with a second universe as the unit of observation, e.g., variables on the NLSY79 Child File.

User Notes: In general, the Center for Human Resource Research does not impute missing values or perform internal consistency checks across waves. Exceptions to this general rule occur when financial support is available, as is the case with the consistency edits performed since 1982 on the NLSY79 fertility data. When bounded interviewing methods are used, responses from the previous interview appear in the text of a question, both to verify that past information and as a point from which to update current information. Bounded interviewing techniques, using data from the *Information Sheets* or flap items (described above), are intended to impose consistency across waves. Data quality checks most often occur in the process of constructing (1) cumulative and current status variables, e.g., 'Highest Grade Completed,' and (2) NLSY79 employment-related variables, e.g., 'Weeks Working in Past Calendar Year,' 'Total Tenure with Employer,' etc. More information on NLSY79 survey instruments can be found in section 3.1.

Reference Numbers

Every variable in the main NLSY79 data files has been assigned a reference number or identifier that determines its relative position within the data file and NLS documentation system. Persons contacting

NLS User Services should be prepared to discuss their question or problem in relationship to the reference number(s) of the variable(s) in question.

NLSY79 (exclusive of the NLSY79 Child File): Once assigned to variables within the NLSY79 data files (exclusive of the NLSY79 Child File), reference numbers remain constant through subsequent revisions of the files. Reference numbers are assigned sequentially, with variables referring to the first survey year having a lower reference number than those variables specific to the second year and so forth. Occasionally variables are created in a year later than that in which the data were actually collected. These variables are frequently given a reference number with a decimal value that reflects the year in which the actual data were gathered rather than the year the created variable was constructed, e.g., R01461.01. Beginning with the 1993 survey, decimals are also used to indicate that more than one variable has been derived from a single question.

User Notes: Reference numbers in the main and geocode data files have traditionally begun with the letter "R." Beginning with the 2000 data release, the work history variables are incorporated with the main data on the same data set. However, these work history variables are assigned reference numbers beginning with "W" for easy identification.

NLSY79 Children: Reference numbers for this data set are the unique identification numbers assigned to each variable that determine its relative position within the codebook. Through 1994, a reference number assigned to a given child variable was specific to a given data release and changed in subsequent releases. All child variables have been arranged in a topical order; variables from the latest release on a given topic are added to those on the same topic from previous years. Different prefixes identified different releases: All variables from the first 1986 release of the NLSY79 child data were preceded by the letter Cxxxxx.xx; second release 1988 variables were preceded by a Dxxxxx.xx, and so forth. Beginning in 1996, all child variables are numbered Cxxxxx.xx.

NLSY79 Young Adults: Reference numbers for this data set are also unique identification numbers assigned to each variable that determine relative position within the codebook. To distinguish them from other NLSY79 variables, young adult reference numbers are prefixed with the letter "Y." These numbers are assigned sequentially to data items in the CAPI questionnaire and the self-report booklet, with decimals indicating that more than one variable has been derived from a single question. Reference numbers for Young Adult variables remain constant through subsequent data releases.

Variable Descriptions or Variable Titles

Each variable within NLSY79 main file data files has been assigned an 80 character summary title that serves as the verbal representation of that variable throughout the hard copy and electronic documentation systems.

Variable titles are assigned by CHRR archivists who endeavor, within the limitations described below, to capture the core CONTENT of the variable and to incorporate within the title (1) AREAS OF INTEREST that facilitate easy identification of comparable variables; (2) UNIVERSE IDENTIFIERS that specify the subset of respondents for which each variable is relevant; and (3) for some variables, REFERENCE PERIODS that indicate the period of time, e.g., survey year or calendar year, to which data refer. Universe identifiers and reference periods are discussed below.

Universe Identifiers: If two ostensibly identical variables differ only in that they refer to different universes, the variable title will include a reference to the applicable universe by either appending in parentheses to each title the appropriate universe (Example 1) or by identifying the universe before the variable title (Example 2).

Example 1: 'Did R Have Any Job since Last Int? (Unemployed or OLF) (1994)', or

Example 2: 'Female - Number of Children R Has Had since Last Interview'

Reference Periods: Variable descriptions may include a phrase indicating the time period to which the data refer. When a date follows a verbal description of a variable and is preceded by the prepositional phrase "in 19XX," the date identifies the calendar year for which the relevant information was collected.

Example: 'Received Income from Child Support in 1991?' This 1992 survey question refers to child support payments received in calendar year 1991.

User Notes: All searches for NLSY79 variables are essentially searches for variable descriptions or titles. Electronic searches of NLSY79 variables via the NLSY79 data set access methods ultimately produce listings of variables by their reference number and variable description or title.

Flexibility in variable title assignment for raw data items is restricted by (1) the actual wording of the question as it appears within the survey instrument; (2) precedent, i.e., how that type of variable has been titled in previous years; and (3) the maximum allowable length for variable titles. An attempt is also made to include key phrases in variable titles so that large groups of variables with similar or related subject matter can be easily identified.

Users should be careful not to presumptively conclude that two variables with the same or similar titles necessarily have the same (1) universe of respondents, (2) coding categories, or (3) time reference period. While the universe identifier and reference period conventions discussed above have been emphasized, users are urged to consult the questionnaires for skip patterns and exact time periods for a given variable and to factor in the relevant fielding period(s) for the cohort.

Variables with similar content, e.g., information on respondents' labor force status, may have completely different titles, depending on the type of variable (raw versus created). In addition, such variables may be located within different NLSY79 areas of interest.

Example 1: 'Employment Status Recode' (ESR), in 1979–98, is the created or reconstructed version of the 'Activity Most of Survey Week' raw variable. The 'Activity' variable is derived from the first question of the full series of questions used by the Department of Labor (DOL) to obtain employment status; the title reflects questionnaire content. ESR, on the other hand, reflects the procedure used to recode the 'Activity' variable. This produces a constructed variable for all respondents based upon responses to the 'Activity' question and all other questions used by the DOL to obtain employment status. These other questions serve to qualify and refine employment status beyond the answer to the initial 'Activity' question.

Example 2: NLSY79 raw fertility variables appear within the various "Children," "Birth Record" or "Birth Record xxxx" areas of interest while edited and constructed versions of these variables appear within the "Fertility and Relationship History/Created" area of interest.

Finally, different archivists, for a period of more than 20 years, have performed the task of assigning variable descriptions to data. While every effort has been made to maintain consistency, users may find some differences in variable title and area of interest assignment.

3.3 NLSY79 Codebook System

All variables present on a main file NLSY79 data set are documented via (1) a codebook; (2) an accompanying codebook supplement; and (3) error updates. This section describes these three primary components of the NLSY79 codebook system and discusses the important types of information found within each. An additional codebook supplement exists for the Geocode data file.

Codebooks

The codebook is the principal element of the NLSY79 documentation system and contains information intended to be complete and self-explanatory for each variable in a data file. The software accompanying the NLSY79 data sets allows easy access to each variable's codebook information and permits the user to print a codebook extract for preselected variables.

Every variable is presented within the NLSY79 documentation as a block of information called a “codeblock.” Each codeblock entry depicts the following important information: reference number, variable title, coding information, frequency distribution, location within the data file, reference to the questionnaire item or source of the variable, and information on the derivation of created variables. Users will find that NLSY79 CAPI codeblocks present greater detail on each variable, including universe totals, universe skip patterns, and range of acceptable values information. Each of these terms is described more completely below. Codeblocks for many variables include special notes containing additional information designed to assist in the accurate use of data from that variable.

Codebooks are arranged in reference number order. Sample reference number assignments used in recent survey years for the NLSY79, NLSY79 Children, and NLSY79 Young Adults are presented in Tables 3.3.1, 3.3.2, and 3.3.3. As a general rule, raw questionnaire items appear first for a given survey year, followed by items from such instruments as the *Information Sheet*, *Employer Supplement*, etc. Variables from the main body of the questionnaire are followed by created or constructed variables drawn from an external data source, such as the *County & City Data Book*. Beginning with the 1993 CAPI surveys, questions relating to each job/employer, which were formerly located within the unique *Employer Supplements*, are merged with the main questionnaire items. A comparison of the reference number assignments used for the 1988 PAPI and 1993 CAPI variables appears in Table 3.3.1; Tables 3.3.2 and 3.3.3 provide users with a sample set of reference numbers for the NLSY79 young adults and children. Users should note that not all survey year assignments will be ordered in precisely this manner, and that the work history reference numbers (Wxxxxx.xx) are not represented here.

Table 3.3.1 NLSY79 1988 & 1993 Reference Number Assignment

1988 PAPI		1993 CAPI	
R25000.–R28927.	All Raw, Edited and Created Variables	R41001.–R44308.	All Raw, Edited and Created Variables
R25000.–R27467.	Questionnaire Items	R41001.–R43988.	Questionnaire Items including the <i>Employer Supplement</i> series
R27469.–R27501.	<i>Information Sheet</i> Items	R43989.–R44036.	<i>Information Sheet</i> Items
R27506.–R27609.	Household Record	R44037.–R44126.	Household Record
R27610.–R28254.	<i>Employer Supplement (ES)</i> ¹		
R28255.–R28371.	<i>Children’s Record Form</i>	R44127.–R44162.	<i>Children’s Record Form</i>
R28372.–R28690.	<i>Childhood Residence Calendar</i> ²		
R28704.–R28729.	Created Variables	R44163.–R44205.	Created Variables
R28735.–R28811.	Supplemental Fertility File Variables		
R28825.–R28927.	Geocode Variables	R44206.–R44308.	Geocode Variables

Note: PAPI refers to paper-and-pencil interviews which were conducted with the NLSY79 during 1979–92. CAPI or computer-assisted personal interviews began for the full NLSY79 cohort in 1993.

¹ Beginning in 1993, variables from the employer supplement series are included within the raw questionnaire items.

² The childhood residence retrospective was unique to 1988 and not refielded.

Table 3.3.2 NLSY79 Young Adult Reference Number Assignment (1994)

Reference Numbers	Contents
Y00001.–Y03836.	All Variables
Y00003.–Y03420.	Questionnaire Items
Y03425.–Y03448.	<i>Children's Record Form</i>
Y03449.–Y03478.	Childcare Roster
Y03479.–Y03556.	Household Record
Y03566.–Y03836.	<i>Self-Report Booklet data</i>

Table 3.3.3 NLSY79 Child Reference Number Assignment

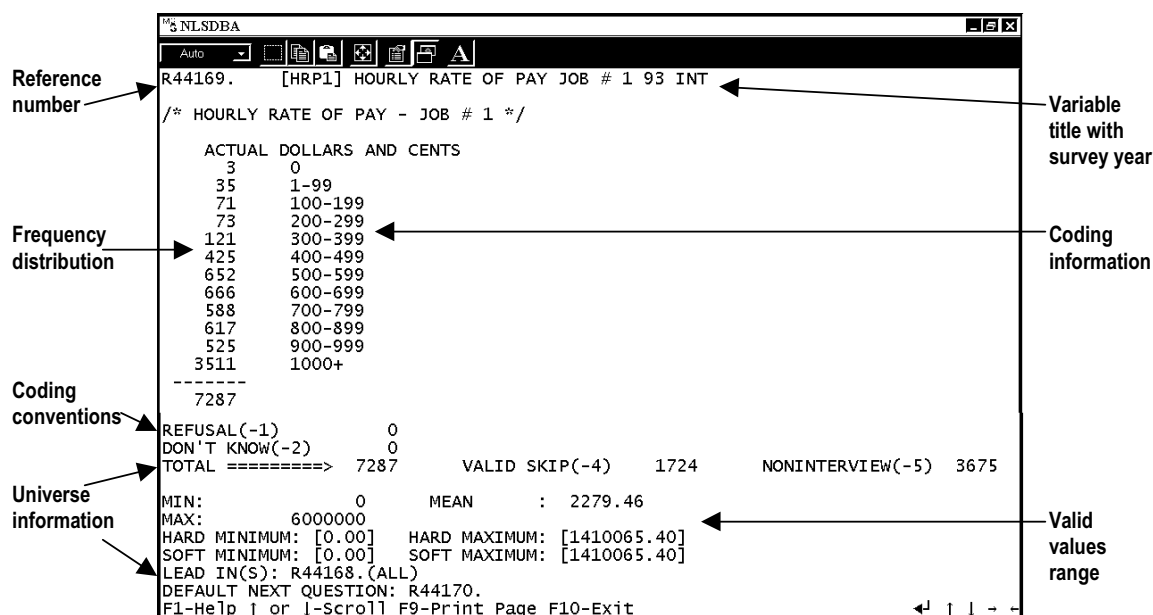
Reference Numbers	Contents
C12000.–C13792.	1994 <i>Child Supplement & Child Self-Administered</i> unedited items
C14000.–C14439.	1994 <i>Mother Supplement</i> unedited items
C14981.–C15089.	1994 raw and normed assessment scores & child sampling weights
C16521.–C17605.	1996 <i>Child Supplement & Child Self-Administered</i> unedited items
C16017.–C16392.	1996 <i>Mother Supplement</i> unedited items
C15637.–C15658.	1996 raw and normed assessment scores & child sampling weights

The following figures give users an example of codebook pages before (Figure 3.3.1) and after (Figure 3.3.2) CAPI implementation.

Figure 3.3.1 NLSY79 Sample PAPI Codeblock

Reference number	(R39562.)	Variable title	HOURLY RATE OF PAY (RATE OTHER THAN HOURLY PREVIOUSLY REPORTED) JOB # 1
	Record type: M92VAR	Question number	Q7773
	Variable name: M92V0387	Survey year	92
Verbatim	HOW MUCH (DO/DID) YOU EARN PER HOUR?		
	ACTUAL DOLLARS AND CENTS		
Frequency distribution	1	1-99	
	4	100-199	
	24	200-299	
	16	300-399	
	112	400-499	
	130	500-599	
	141	600-699	
	141	700-799	
	128	800-899	
	108	900-999	
	591	1000+	
Coding conventions	REFUSALS(-1) 46	DON'T KNOW(-2) 56	INVALID SKIPS(-3) 16
Universe information	VALID SKIPS(-4) 7502	NONINTERVIEWS(-5) 3670	
	MINIMUM: 41	MAXIMUM: 15000	

Figure 3.3.2 NLSY79 Sample CAPI Codeblock



Coding Information: Each codeblock entry presents the set of legitimate codes that a variable may assume along with a text entry describing the codes. *Users should note that coding information for a given variable in the NLSY79 codeblock is (1) not necessarily consistent with the codes found within the questionnaire and (2) not necessarily consistent for the same variable across years. Use only the codebook coding information for analysis.* The following types of code entries occur in NLSY79 codeblocks:

Dichotomous (or variables answered yes/no) are uniformly coded “Yes” = 1, “No” = 0. Other dichotomous variables have frequently been reformulated so this convention may be followed.

Discrete (Categorical), as in the case of the NLSY79 example, the variable ‘Activity Most of Survey Week CPS Item’.

- | | |
|---------------------------|-------------------|
| 1 WORKING | 5 GOING TO SCHOOL |
| 2 WITH A JOB, NOT AT WORK | 6 UNABLE TO WORK |
| 3 LOOKING FOR WORK | 7 OTHER |
| 4 KEEPING HOUSE | |

Continuous (Quantitative), as in the case of hourly rate of pay in the example above. These variables have continuous data but are presented in the codebook using a convenient frequency distribution. NLSY79 users will note that most valid data are positive numbers. Special cases are flagged by negative numbers in the NLSY79 including NLSY79 Children. See Appendix 13 in the *NLSY79 Codebook Supplement* for more detail on the handling of negative numbers in the data files. The following conventions have been used throughout the data:

Noninterview	-5
Valid Skip	-4
Invalid Skip	-3
Don't Know	-2
Refusal	-1

Derivations: The decision rules employed in the creation of main file constructed variables have been included, whenever possible, in the codebook under the title “DERIVATIONS.” This information enables researchers to determine whether available constructs are appropriate to their needs. In the case of the example NLSY79 variable in Figure 3.3.1, no derivation is shown because these variables are picked up directly from the interview schedule. Certain variables will contain a reference to an appendix for the decision rules that were used in creating the variable.

Frequency Distribution: In the case of discrete (categorical) variables, frequency counts are normally shown in the first column to the left of the code categories. In the case of continuous (quantitative) variables, a distribution of the variable is presented using a convenient class interval. The format of these distributions varies.

Questionnaire Item: “Questionnaire item” is a generic term identifying the printed source of data for a given variable. A questionnaire item may be a question, a check item, or an interviewer’s reference item appearing within one of the survey instruments.

The questionnaire location for NLSY79 entries appears either in parentheses or brackets directly after the reference number, for example R04434. (SO6D1314). The five questionnaire item numbering conventions used in the codebook are described in the “Survey Instruments” section of this chapter (see especially Table 3.1.2).

Before the adoption of CAPI if an NLSY79 variable was not taken directly from one of the survey instruments, the questionnaire location contained an asterisk (*) in the codebook. The following categories of variables had no questionnaire numbers: (1) assigned identification numbers for the respondent, child, or family unit, etc.; (2) all derived or constructed variables; (3) variables from the following special surveys: Profiles (*ASVAB*), the School Survey, and the Transcript Survey; (4) variables found on constructed data files such as the Supplemental Fertility File (area of interest “Fertility and Relationship History/Created”); and (5) variables drawn from an external data source such as those found on the geocode files. In CAPI years, survey staff assign a question name that is not used in the questionnaire. This name remains the same in subsequent rounds, so similar created variables can be easily located.

Section, deck, and question numbers have been somewhat arbitrarily assigned by NORC to the information and questions found in special survey instruments such as the *Household Screener*, *Information Sheet*, *Children's Record Forms*, *Household Interview Forms*, and the *Employer Supplements*. The section and deck numbers for these special survey items were numbered sequentially after the main survey items and their specific order varies each year. The exception to this is the assignment of the deck numbers for the *Employer Supplements*. Question numbering is discussed earlier in the "Survey Instruments" section of this guide (see especially Table 3.1.3).

Variables created or modified for the child data set contain a mnemonic variable name in the source number location. NLSY79 Young Adult questions are numbered like the NLSY79 main questions (see Table 3.1.2); questions begin with Q and are followed by section number, question number, and loop. Question (source) numbering is discussed in more detail in section 3.1 of this guide.

Universe Information: Universe information was attached to select 1979–92 variables. Beginning with the 1993 CAPI interviews, the amount of universe information was expanded to include:

1. **Universe Totals:** Two totals are presented: (1) the sum of the frequency counts for each coding category is presented below the individual codes; and (2) the sum of the valid responses plus missing response counts of "refusals," "don't knows," and "invalid skips" can be found in the TOTAL===== field. The number of respondents who legitimately did not respond to a question, i.e., "valid skips (-4)" and "noninterviews (-5)," are also depicted.
2. **Universe Skip Patterns:** The following detailed universe information will enable researchers to easily trace the flow of respondents both backward and forward through various parts of the CAPI questionnaire items included in the codebook:

"Go to Reference # XXXXX," appended to certain coding categories, indicates that respondents selecting that answer category were routed to the next question specified.

"Lead In(s) Reference # XXXXX." identifies the question or questions immediately preceding the codeblock question through which the universe of respondents was routed. Each lead-in reference number is followed by the relevant response value indicators, e.g., (Default), (ALL), [1:1], [1:6], etc. For example:

- | | |
|----------------------|--|
| R41000. (All) | This means that all cases where R41000. is asked will branch to the current question. This does not imply all respondents are asked question R41000. |
| R41000.
(Default) | This means that the default path of control from question R41000. is to branch to the current question, but there may be conditions under which a different path would be taken. |
| R41000. [1:6] | This means that whenever the response category for question R41000. takes on the values one to six inclusive, the next question is the current question record. |

"Default Next Question" specifies the next question that all respondents of the current codeblock will be asked unless some other skip condition indicates otherwise.

Valid Values Range: Depicted below the frequency distribution is information relating to the range of valid values for that particular distribution. “MINIMUM” indicates the smallest recorded value exclusive of “NA” and “DK.” “MAXIMUM” indicates the largest recorded value. The computer-assisted interview contains internal range checks that limit responses to those between predesignated values, alert interviewers to verify non-normative values, and bolster the information provided by the traditional minimum and maximum fields (see, for example, Figure 3.3.2).

Maximum and Minimum Fields: The MIN and MAX fields define the range, i.e., the lower limit and the upper limit, of data values for a given question. A MAX of \$156,359 on an income question, for example, means that this value was the highest value recorded.

Hardmax and Hardmin Fields: Hard Maximum and Hard Minimum fields denote the highest and lowest values that were accepted by the CAPI program. A Hardmax of 500,000 and a Hardmin of 0 on an income question indicate that no values above \$500,000 or values lower than zero (no income) can be accepted. Dates, e.g., month/day/year of the respondent’s last interview [lintdate] and current interview [curdate], are used as Hardmin and Hardmax values in order to restrict responses to certain questions to values within that range. Responses outside this range must be entered by the interviewer in the comment field.

Softmax and Softmin Fields: Softmax and Softmin fields cover ranges where an answer may exceed reasonable limits yet remain within the absolute limits and are acceptable after verification. A Softmax set to \$80,000 on an income question will cause the machine to “beep” and a warning to appear on the screen. Interviewers are thus alerted that the value is unusual and the respondent’s answer should be verified.

Restricted Income Values: Confidentiality issues restrict release of all income and asset values. To insure respondent confidentiality, the values of income or asset variables exceeding particular limits are truncated and the upper limits converted to a set maximum value. From 1979 through 1984, the upper limit on income variables was \$75,000, and any amounts exceeding \$75,000 were converted to \$75,001. Beginning in 1985, the upper limit on income amounts was increased to \$100,000 due to inflation and the advancing age of the cohort, and amounts exceeding \$100,000 were converted to \$100,001. During that same survey year, specific questions on assets owned by NLSY79 respondents were added to the survey.

The asset amounts have different upper limits, and the types of variables and limits for those variables are as follows: (1) mortgage, market value, and debt on residential property and total market value of assets each worth more than \$500 and miscellaneous debt more than \$500—more than \$150,000 converted to

\$150,001; (2) market value and debt on a farm or business and savings—more than \$500,000 converted to \$500,001; (3) market value and debt on vehicles—more than \$30,000 converted to \$30,001. Beginning in 1989, the amounts exceeding the upper limits mentioned above were assigned the average value of all values exceeding the limits, in an effort to more accurately reflect the true range of income and asset values.

In the unique instance where one case has a value above the 1985 truncation limit, the value for that case is assigned the truncation limit. Finally, beginning in 1996, the top two percent of respondents with valid values were averaged and that average value replaced all values in the top range. Users should be aware of these changes in the income ceiling if they are carrying out longitudinal analyses with these data. Upward trends in mean income statistics may reflect this change in the ceiling value. More information about truncation is available in the “Income” and “Assets” sections of this guide.

Verbatim: Generally during the PAPI years, when a NLSY79 variable was taken directly from the questionnaire, the verbatim of the question appeared beneath the variable title. If a question is the source for more than one variable, the first variable contains the verbatim while subsequent variables prompt the user to refer back to the variable containing the verbatim. The following verbatims appear for reference numbers R03194. and R03195. and demonstrate this convention.

R03194. ‘In Which Months of 1979 Did You (or Your Husband/Wife) Receive Supplemental Security Income? January 80 INT’

R03195. ‘See R (3194.) February’

Codebook Supplements

There are two NLSY79 codebook supplements. The first supplement contains variable creation procedures, supplementary coding categories, and derivations for selected variables on the main NLSY79 data files. Information provided within this document is **not** available in the hard copy NLSY79 codebooks, nor will it be found on the electronic documentation files on the NLSY79 data sets. The other supplement contains comparable information specific to the NLSY79 geocode data files.

Although there is no separate codebook supplement for the NLSY79 Child File, references will be found within the codeblock of select child variables to NLSY79 main file attachments and appendices. There is also a booklet containing several young adult attachments. A paper version of the 2002 young adult survey containing several attachments similar to the main cohort *Codebook Supplement* is also available. A brief description of the contents of the *NLSY79 Codebook Supplement* and the young adult attachments appears at the end of this section. Child data users are encouraged to consult these documents.

Attachments & Appendices: NLSY79 Main File Codebook Supplement

Attachment 3: Industry and Occupation Codes is a compilation of (1) the 3-digit 1970 Census classifications used to code job and training information as well as occupational aspiration information and *Employer Supplements* (U.S. Census Bureau, “1970 Census of Population Alphabetical Index of Industries and Occupations,” U.S. Government Printing Office, Washington, DC, 1971); (2) the 3-digit 1980 Census codes that have been used in addition to the 1970 codes, beginning with the 1982 survey, to classify the industry and occupation of respondents’ most current or most recent job (CPS job) (U.S. Census Bureau, “1980 Census of Population Alphabetical Index of Industries and Occupations,” U.S. Government Printing Office, Washington, DC, 1981); and (3) the 1977 military occupational specialty codes used to classify responses to the 1979–85 questions on military jobs and military occupations (U.S. Department of Defense, “Occupational Conversion Manual: Enlisted/Officer/Civilian,” Defense Manpower Center, Arlington, Virginia, DOD 1312.1-M).

Attachment 4: Fields of Study in College provides the coding classifications for: (1) the 1979–83 major field of study at current or last college attended and (2) the 1984–2002 major field of study at most recent colleges attended.

Attachment 5: Index of Labor Unions and Employee Associations provides codes for the 1979 questions on name of union/employee association at jobs #1 – #5 (i.e., R00937.–R00941.).

Attachment 6: Other Kinds of Training lists the various categories of occupational training used to code the 1979 survey question on types of other training programs in which a respondent was enrolled for at least one month (R01348., R01353., R01358., R01363.).

Attachment 7: Other Certificate Codes defines codes for the various types of certifications (i.e., practical nurse, welding, insurance, chef, etc.) that a respondent had ever received as of the 1979 interview (R01376., R01377., R01378., R01379.).

Attachment 8: Health Codes provides a modified version of the International Classification of Diseases (ICD-9) codes [*International Classification of Diseases, Volumes 1 & 2*. Geneva, WHO, 1977–78], which were used to classify types of health problems limiting the amount or kind of work a respondent could do during survey years 1979–82 and the work-related injury data collected during the 1988–90 and 1992–2002 surveys. Also included is a list of numeric codes identifying the parts of the body affected by health problems.

Attachment 100: Geographic Regions provides a listing of those states which comprise each of the four regions used in such variables as ‘Region of Residence’, ‘South/Non-South Place of Birth’, and ‘South/Non-South Place of Residence at Age 14’.

Attachment 101: Country Codes provides the foreign country codes used to identify the respondent’s country of residence, country of parents’ birthplace, and country of citizenship at time of immigration.

Attachment 102: State ‘Federal Information Processing Standards’ or FIPS Codes lists the codes which were used for respondent’s state of birth and state of residence.

Attachment 103: Religion Codes contains the various denomination categories used to code the 1979 religion of respondent questions (R00103.10 and R00104.10) and the 1982 religion questions (R06558., R06583., R06586., R06613., and R06616.).

Attachment 106: *Profile of American Youth* provides general and technical information on the 1980 administration of the *ASVAB (Armed Services Vocational Aptitude Battery)* to NLSY79 respondents. Included in this attachment are technical notes on the *ASVAB* scale scores, an annotated bibliography of DOD publications, an example of the test score report, and various brochures disseminated to participating respondents. An *Addendum* provides information on the creation of two Armed Forces Qualifications Test scores, AFQT80 and AFQT89, which were added to the data set beginning with the 1979–90 release. Note: This attachment is **not** within the codebook supplements; it is a separate document.

Appendix 1: Employment Status Recode (ESR) Variable Creation 1979–1998 contains the adapted version of the FORTRAN program and subsequent SPSS program used to create this measure of main labor force activity during the survey week. This variable was not created in 2000 or 2002.

Appendix 2: Total Net Family Income Variable Creation 1979–2002 provides the code used to create this *KEY* income variable, as well as the poverty level and poverty status variables for each survey year.

Appendix 3: Job Satisfaction Measures 1979–1982 provides background information and yearly reference numbers for both the scale items and global satisfaction measures of the modified Quality of Employment Survey scale administered in the 1979–82 surveys. Additional references and a methodology for constructing the full scale are also presented.

Appendix 4: Job Characteristic Index 1979 & 1982 provides background information, reference numbers, questionnaire locations, and additional references for the job complexity questions asked in these two survey years.

Appendix 5: Supplemental Fertility File Variables 2002 provides (1) a brief overview of the contents of the 1979–2002 “Fertility and Relationship History/Created” area of interest on the main NLSY79 data, (2) background information on the 1982 data quality check, (3) background on the 1994 data reconciliation, and (4) the availability of additional reports assessing NLSY79 fertility data.

Appendix 6: SMSA Urban-Rural Creation contains the decision rules used to create (1) the four codes (not in SMSA, SMSA not central city, SMSA central city not known, and SMSA central city) for the ‘Current Residence in SMSA’ variables and (2) the urban and rural codes for the ‘Is R’s Current Residence Urban/Rural?’ variable series.

Appendix 7: Unemployment Rate provides an explanation of how the variable ‘Unemployment Rate of Labor Market of Current Residence’ was created.

Appendix 8: Highest Grade Completed and Enrollment Status Variable Creation: 1990–2002 contains the codes used to create the *KEY* 1990–2002 variables ‘Highest Grade Completed as of May 1 Survey Year’ and ‘Enrollment Status as of May 1 Survey Year.’

Appendix 9: Linking Jobs through Survey Years identifies the procedures and variables necessary for linking employers reported across contiguous survey years.

Appendix 11: NLSY79 Round 12 (1990) Survey Administration Methods describes the 1990 experiment with PAPI versus CAPI methods of interviewing.

Appendix 12: Most Important Job Learning Activities - 1993–1994 provides variable reference numbers, titles, and value labels for the four 1993 and 1994 items which identify method(s) used by respondents in learning to perform job duties associated with their current or most recent job.

Appendix 13: Introduction to 1993 through 2002 CAPI Questionnaire and Codebook discusses the changes caused by moving from paper-and-pencil interviewing (PAPI) to computer assisted personal interviewing (CAPI). The appendix discusses changes and new documentation items, new terms, coding convention changes, data and codebook reordering, and changes in data collection procedures.

Appendix 14: 1993–2002 Instrument Rosters describes the selected rosters, or matrices of data, that are constructed during a CAPI interview. Rosters are created so that interviewers have a complete table of information to either update or to show to a respondent so he/she may choose an answer. Examples of rosters are CHILD, EMPLOYER, and HOUSEHOLD. The CHILD data matrix provides interviewers with information such as ID, gender, and birth date. The EMPLOYER matrix provides name, start and stop work dates, and whether the respondent is still working for the employer. The HOUSEHOLD

roster lists the gender, age, highest grade completed, and relationship to the youth of all members in the household.

Appendix 15: Reciprocity Event Histories describes how data are collected for NLSY79 respondents who receive government assistance such as AFDC, Food Stamps, Unemployment Compensation, and other programs. Reciprocity data are described for both PAPI and CAPI interviews.

Appendix 16: 1994 Recall Experiment describes a special test run during the 1994 NLSY79 survey. To measure the effects of switching from an annual survey to a biennial survey, a recall experiment was implemented. Approximately 10 percent of NLSY79 respondents were asked to recall data over a two-year period. Respondent answers to questions covering a two-year period rather than a typical one-year period were then compared to the answers given during the 1993 survey to understand the biases that would result from skipping a year of interviewing.

Appendix 17: Interviewer Characteristics Data describes the data available regarding the characteristics of NLSY79 interviewers; this information is based on NORC's interviewer personnel files. This new data, starting with the 1996 survey, permit researchers to link interviewers with the respondents they interviewed.

Appendix 18: Work History Data explains the programs used to create the work history arrays. Also provided is a list of items used as inputs to the work history programs and information on the coding of selected variables.

Appendix 19: SF-12 Health Scale Scoring discusses the source of the SF-12 scale, part of the over-40 health module described in the "Health" section of this guide, and overviews the creation of scale scores for NLSY79 respondents.

NLSY High School Transcript Surveys: Overview and Documentation contains background information on the sample design, field work, and types of variables collected during the three rounds of this special survey. Included is a transcript survey codebook, instructions for coding courses, course codes, and copies of the transcript coding form and school questionnaire, as well as additional references to other technical reports prepared by the sponsoring agency, the National Center for Research in Vocational Education. Note: This document is separate from the codebook supplements.

Attachments & Appendices: NLSY79 Geocode Data File Codebook Supplement

Appendix 10: Geocode Documentation provides year by year descriptions of how the geocode files were constructed, important information on changes in SMSA designations, and detailed explanations of the missing values for the geocode variables.

Attachment 100: Geographic Regions provides a listing of those states which comprise each of the four regions used in such variables as 'Region of Residence' and 'South/non-South Place of Birth/Place of Residence at Age 14.'

Attachment 101: Country Codes provides the foreign country codes used to code respondents' country of residence and country of parents' birthplace.

Attachment 102: State 'Federal Information Processing Standards' or FIPS Codes (U.S. Department of Commerce, National Bureau of Standards) lists the codes used for respondents' state of birth and state of residence.

Attachment 104: SMSA Codes contains the coding information used to classify SMSA, MSA, CMSA, and PMSA of residence at each interview date.

Attachment 105: Addendum to FICE Codes contains the supplementary identification numbers for those colleges and universities not listed in the *Education Directory-Colleges and Universities* (1981–82 and 1982–83 Supplement) published by the National Center for Education Statistics. It also contains detailed information on the revised FICE code series (FICE codes and FICE types).

Appendix 7: Unemployment Rate provides an explanation of how the continuous and collapsed versions of the variable 'Unemployment Rate for Labor Market of Current Residence' were created.

Attachments & Appendices: NLSY79 Young Adult Attachments

Beginning in 2000, these documents are part of the printed questionnaires.

Attachment 3: 1970 Census Industry/Occupation Codes lists the 3-digit 1970 Census classifications used to code job and training information and *Employer Supplements* (U.S. Census Bureau, "1970 Census of Population Alphabetical Index of Industries and Occupations," U.S. Government Printing Office, Washington, DC, 1971).

Attachment 4: 1990 Census Industry/Occupation Codes lists the 3-digit 1990 Census classifications used for double coding of occupation and industry for the CPS job (U.S. Census Bureau, "1990 Census of

Population Alphabetical Index of Industries and Occupations,” U.S. Government Printing Office, Washington, DC, 1991).

Attachment 5: Q by Q Young Adult Specifications reproduces the instructions provided to interviewers for the administration of specific questions throughout the questionnaire. This attachment is intended to permit researchers to determine what types of help and information were available to interviewers during the survey and is comparable to the interviewer reference manuals for the main NLSY79.

Error Updates

Prior to working with an NLSY79 data file, users should make every effort to acquire information on current data and/or documentation errors. A variety of methods are used to notify users of errors in the data files and/or documentation and to provide those persons who acquired a NLSY79 data set from the Center for Human Resource Research with corrected information. Errors discovered after the release of a data file are distributed in hard copy form to current data purchasers along with the data set. Errata can be accessed online from the NLS web site at <http://www.bls.gov/nls> by following links for the cohort of interest. Error notices and information on how to acquire the corrected data and/or documentation also appear in *NLS News*, the quarterly NLS newsletter, which is available online at <http://www.bls.gov/nls/>.

3.4 Data Set Search Functions

NLSY79 variables can be accessed via areas of interest or through a search of variable titles for any word. Both search functions provide users with bridging information to the codebook and survey instruments.

Areas of interest. NLSY79 data files are organized so that variables sharing a common factor are stored in unique groupings called “areas of interest.” Users can browse through a given area and examine the variables associated with that topic. NLSY79 areas of interest are listed in Appendix C of this guide.

Any word search. All words, numbers, and symbols found in any variable title form an index on the data set. The “Any Word in Context” search function in the search software allows the user to search this index and select NLSY79 variables whose titles contain any single word or combination of words. Every word, number and symbol found in each variable title has been used to form a dictionary or index and can be used to conduct a search. This function broadens the ability of the user to access variables on a given topic but is still dependent on the wording of each variable title, which in turn is questionnaire-dependent.

User Notes: Any “Word in Context” searches for NLSY79 variables are limited by the choice of variable titles. Flexibility in variable title assignment for raw data items is restricted by the wording of the question as it appears in the survey instrument and the maximum allowable length for variable titles.

Accessing Variables by Area of Interest

NLSY79 data files are organized in such a way that variables sharing a common factor such as longitudinality, topic, research use, or source are assigned an “area of interest.” This section (1) discusses the decision rules used to assign a variable to a given area of interest and (2) describes the hard copy presentations and electronic search functions available to help users access NLSY79 variables by area of interest. Originally, the primary function of the NLSY79 area of interest file structure was to provide magnetic tape users with the physical location of a variable for extracting purposes. Today it is used to assist data set users in locating variables on a given topic.

Record Types: The 1979–93 NLSY79 data releases were structured on the basis of three types of records. These record types also denoted the file structure of the data files. The decision rules for assigning variables to these records were as follows:

1. ***Longitudinal or Repeating:*** Questions that were asked or variables that were created in an identical manner in multiple survey years were placed within longitudinal or repeating records. Examples of these record types were: Key Variables, CPS, Jobs, Job Information, Periods Not Working Within Job Tenure, School, Income, Assets, Child Record Form/Biological, Household Record, Between Job Gaps, Government Jobs, Military, Interviewer Remarks.
2. ***Topical or Non-Repeating:*** Variables that shared some common research topic were grouped together into subject-related record types. These included Marriage, Degrees & Certificates, Health, Childcare, Alcohol, Drugs, Profiles, Geocode xxxx (Geocode CD only), Birth Record xxxx, Family Background, Last Interview Information, Children, School Survey, Transcript Survey, Training, Illegal, Attitudes & Influence, Attitude, Fertility and Relationship History/Created, and Time Use. These records contained topically related questions, regardless of their recurring status. Inclusion in such a record type did not exclude the possibility that a given topical area of interest contained any longitudinal data, or that longitudinal data on a given topic was available elsewhere within the data set as a whole.
3. ***Miscellaneous:*** All other non-longitudinal variables, i.e., those questions that had been asked or variables that had been created only in select survey years, were placed within year-specific Misc. xxxx

or miscellaneous records. These generally represented groupings of unrelated sets of variables based on questions that had not been asked in a consistent manner over a significant number of years.

Beginning with the 1994 release, the data ceased to be released on magnetic tapes. The records described above no longer reflect the physical structure of the data file, but became a search tool by which to classify data items. With the 1998 release, the reference to “records” was abandoned in both documentation and software applications in favor of “areas of interest” to more accurately reflect the less restrictive topical nature these classifications now have. There has been a good deal of consistency in the assignment of data items to specific areas of interest, many of which are named identically to the historical record types. However, some traditionally longitudinal areas of interest now contain some data items that are not necessarily repeating. And new areas of interest can be designated more easily, to reflect the variables from a particular source or of a specific topical interest. In addition, the names of some areas of interest were lengthened to be more understandable. Old and new names are listed in Appendix D.

Recurring variables can often be found by examining the list of variables in traditionally longitudinal areas of interest. Many of these are also found in areas of interest that are assigned largely on a sectional basis such as “Marriage,” “Health,” etc., where the sections are repeated in multiple surveys. Users should note that while the variables are grouped under topical names, every item relating to a particular subject area will not necessarily be found in the area of interest with the generic name. For example, while the majority of main youth childcare variables can be found in the area of interest “Childcare,” other areas of interest, such as “Child Record Form/Biological,” “Government Jobs,” “Government Training,” “Time Use,” and “Misc. xxxx,” also contain variables which may be of interest to those focusing on childcare issues.

User Notes: Once placed within an area of interest, variables are seldom moved. However, there are certain exceptions to this general rule. Beginning with the 1988 release, several sets of NLSY79 main file variables dealing with alcohol use, government training, and other training were deleted from the “Misc. xxxx” areas of interest and reassigned to the “Alcohol,” “Government Training,” or “Training” areas of interest. However, certain other variables from the “Government Training” and “Other Training” sections of the 1979 and 1982 surveys were not moved; R01368. through R01374. remain in “Misc. 1979”; R01375. through R01404. remain in “Degrees & Certificates”; and R07443., R08281., and R08282. remain in “Misc. 1982.” Users should be aware that while variables placed in longitudinal or “repeating” areas of interest are generally present for all survey years, some variables will not be found there for some years due to discontinuation or a change in the form of the question or series of questions. Likewise, although variables placed in the miscellaneous areas of interest will not necessarily have been asked in a consistent manner in all years present, they may exist in similar form for more than just one or two years—possibly quite a few.

NLSY79 Child Data Areas of Interest

The NLSY79 1979–2002 child data on data file have been grouped into areas of interest reflecting the topical and instrument-related organization of the data file itself. For complete information on child data areas of interest consult the *NLSY79 Child & Young Adult Data Users Guide*.